

The VIMSS Comparative Genomics Website

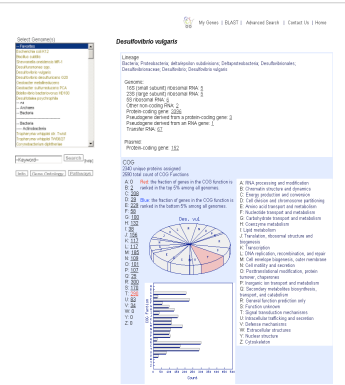
Katherine Huang¹, Morgan Price¹, Richie Koche², Keith Keller², Eric Alm¹, Adam Arkin^{1,2}

¹Division of Physical Sciences, Lawrence Berkeley National Lab; ²Department of Bioengineering and Chemistry, University of California, Berkeley



Abstract

The VIMSS microbial comparative genomics website is designed specifically to facilitate multi-species comparison. The highlights include the operon and regulon predictions, the multi-species genome browser, KEGG metabolic pathway browser, the VertiGO gene ontology browser and the Bioinformatics Workbench. Personalized lists of genes from the website can be identified and deposited in the gene carts for further analysis, such as multiple sequence alignment and phylogenetic trees. Finally, our web tools allow users to add their own annotations to genes that become available via our website. The VIMSS comparative genomics website and tools are freely available to the science community.



Genome Menu

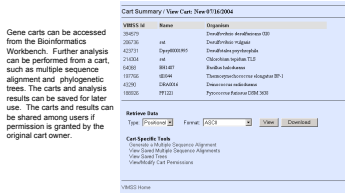
Currently, the VIMSS genome database contains 161 complete microbial genomes. The genome menu, displayed as the left side bar on all VIMSS web pages, is designed so that one or more species or phylogenetic groups can be selected for further navigation. Frequently inquired genomes, such as *E. coli*, *K12*, *S. aureus*, and the GTL genomes, are conveniently grouped into "Favorites", which is provided on the top of the genome menu.

Genome Page

The genome page offers detailed information on each genome, including the full phylogenetic lineage, a summary of the genomic and plasmid tool and an overall profile of the COG (Cluster of Orthologous Groups²) function categories. The COG functions are color-coded in red or blue if the fraction of genes in that category is ranked in the top or bottom 5% of all microbial genomes.

Bioinformatics Workbench

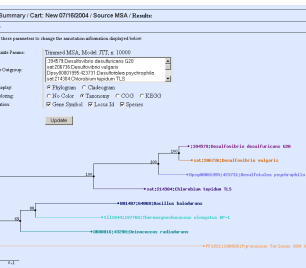
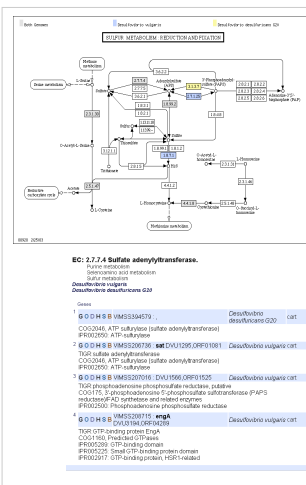
Users can easily manage genes of interest by depositing them in the personalized and password protected gene carts. Gene information and sequences can be retrieved and downloaded from the carts.



Metabolic Pathway Viewer

The metabolic pathway viewer uses the metabolic maps obtained from KEGG (Kyoto Encyclopedia of Genes and Genomes³). The metabolic pathways are mapped out based on whether the enzymes are found in the genomes. The users can select two genomes and view the differences in the pathways. The screenshot of the sulfur metabolism map of *D. vulgaris* and *D. desulfuricans* G20 is shown below. Enzymes that are found in both genomes are colored in gray; those found only in *D. vulgaris* are shown in blue and those found only in *D. desulfuricans* G20 are shown in yellow. Enzymes that are not found in either genome are shown in white.

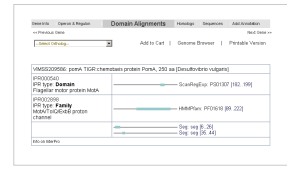
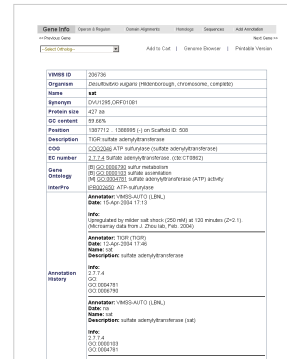
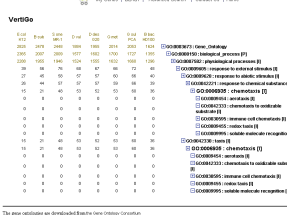
Each feature in the metabolic map is clickable, such as links to other maps and enzymes. When an enzyme is found in a genome, the enzyme is hyperlinked to a list of all genes that are assigned to that enzyme.



VertiGO

VertiGO is a gene ontology (GO) browser developed by the VIMSS computational core, which allows the users to select multiple genomes from the genome menu to browse the GO hierarchy. The GO hierarchy was developed by the Gene Ontology⁴ consortium⁵. It provides a means of organizing all the gene annotations in a genome.

As shown in the screen shot below, 8 genomes are chosen, and the number of unique genes of each GO term is displayed for each genome. *E. coli* K12 has total 2825 genes that have GO terms assigned to them, whereas *D. vulgaris* has 1884. The VertiGO browser can compare several genomes at once. For instance, *D. vulgaris* has 53 out of 1884 genes that are linked to chemotaxis, which is a lot more than *E. coli* that has 15 out of 2825.



Genome Browser

The VIMSS Genome Browser allows viewing and aligning multiple genomes. The selected genomes are aligned by an anchor gene and all genes within the view are color-coded by the putative orthology relationships.

Users can add or delete selected genomes, zoom in or out, or scan upstream or downstream. Each gene (feature) on the browser is a hyperlink and clicking on a gene can take one of three actions: loading the locus info page, choosing a different gene anchor or adding that gene to a gene cart.

Locus Info

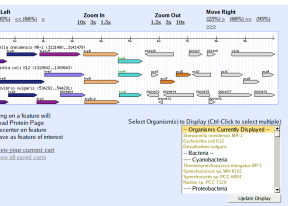
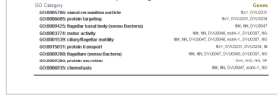
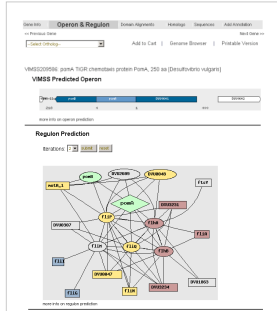
The locus info page provides detailed information on each locus. The page is divided into 6 sections: gene info, operon and regulon, domain alignments, homology, sequences and annotation page.

Gene Info: A table that shows a summary of each locus, such as gene name, gene position, COG, EC, domains, gene ontologies and annotation history.

Operon and Regulon: At VIMSS operon predictions are displayed graphically on the website as part of the locus info page. Experimentally characterized operons are also provided when available.

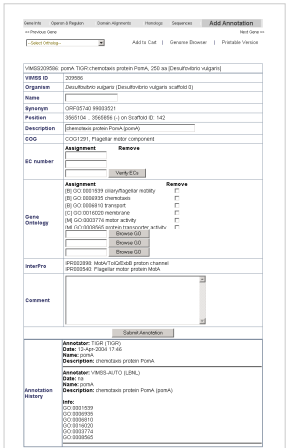
In the VIMSS Operon Viewer, the query gene is shown in light blue. Nearby genes predicted to be in the same operon are colored in dark blue. The next genes upstream and downstream of the operon are shown in white. The arrowhead indicates the transcription direction. The numbers appear below the genes indicate the intergenic distance between two adjacent genes. Negative numbers indicate overlapping genes. Each gene is labeled by its gene name, synonym or VIMSS identifier, and is hyperlinked to its locus info page.

A regulon is depicted as a graph of interconnected genes, and the query gene is diamond-shaped. Blue edges connect genes with similar expression profiles (microarray data), black edges connect genes that are co-localized in several distantly related genomes, and red edges connect genes when both methods are in agreement. Genes in the same operon are indicated by the same color, other genes are shown in gray. Genes that are directly connected to the query gene are oval-shaped, and genes two links out are square-shaped. More distant connections are available using the pull-down menu.



Annotation

The annotation form is provided as part of the locus info page. Users may assign gene name, add or delete EC (enzyme commission) numbers, add or delete GO terms, and add comments. Once the annotation is submitted, it immediately appears in the annotation history. Annotations are propagated into the VIMSS genome database once verified.



References

1. Tatufove RL, et al. A genomic perspective on protein families. Science. 1997; 278:383-387.
2. Tatufove RL, et al. The COG database: an updated version includes eukaryotes. BMC Bioinformatics. 2003; 4(1):41.
3. www.kegg.org
4. www.geneontology.org

Acknowledgment

The VIMSS comparative genomics website is sponsored by the DOE GTL grant, DE-AC03-78SF0008.

<http://vimss.org>